

Facial Expression Recognition using Firefly-based Feature Optimization

Kamlesh Mistry¹, Li Zhang² and Graham Sexton²

¹School of Computing, Teesside University, UK

²Department of Computer Science and Digital Technologies, Faculty of Engineering and Environment, Northumbria University, Newcastle, UK, NE1 8ST

Yifeng Zeng and Mengda He

School of Computing
Teesside University
Middlesbrough, UK

Abstract—Automatic facial expression recognition plays an important role in various application domains such as medical imaging, surveillance and human-robot interaction. This research proposes a novel facial expression recognition system with modified Local Gabor Binary Patterns (LGBP) for feature extraction and a firefly algorithm (FA) variant for feature optimization. First of all, in order to deal with illumination changes, scaling differences and rotation variations, we propose an extended overlap LGBP to extract initial discriminative facial features. Then a modified FA is proposed to reduce the dimensionality of the extracted facial features. This FA variant employs Gaussian, Cauchy and Levy distributions to further mutate the best solution identified by the FA to increase exploration in the search space to avoid premature convergence. The overall system is evaluated using three facial expression databases (i.e. CK+, MMI, and JAFFE). The proposed system outperforms other heuristic search algorithms such as Genetic Algorithm and Particle Swarm Optimization and other existing state-of-the-art facial expression recognition research, significantly.

Keywords—feature selection, facial expression recognition, and firefly optimization.

I. INTRODUCTION

Automatic facial expression recognition has become a new hotspot of AI research and shows great potential in benefiting a wide variety of applications, e.g. personalized healthcare [1], interactive video games [2], human robot interaction [3, 4] and surveillance systems [5]. However, it is still a difficult task to select significant discriminating facial features that could represent the characteristics of each expression because of the subtlety and variation of facial expressions.

In order to deal with the above challenge, this research proposes a facial expression recognition system with a modified Local Gabor Binary Patterns (LGBP) for discriminative feature extraction and a modified variant of firefly algorithm (FA) for feature optimization. In order to overcome illumination changes, rotation and scale variations, first of all, an extended overlap LGBP operator is proposed to generate an initial refined facial representation for an input image. Then the FA variant is proposed to reduce feature dimension and identify the most significant discriminative facial features. In order to mitigate premature convergence of the conventional FA, this FA variant employs Gaussian, Cauchy and Levy distributions to further mutate the most

promising solution identified by the FA to enable global exploration in the search space. Finally, multiple classifiers, such as Artificial neural networks (NN), Support Vector Machine (SVM), NN-based ensemble and SVM-based ensemble, are employed to recognize seven emotions including, anger, disgust, happiness, sadness, fear, surprise and neutral. Evaluated with the extended Cohn-Kanade (CK+) [6], JAFFE [7], and MMI [8] databases, the proposed system outperforms conventional optimization algorithms such as FA, Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and other state-of-the-art facial expression recognition research, significantly. The overall system architecture is shown in Figure 1.

The rest of the paper is organized as follows. Section II presents related research. Section III introduces the proposed facial expression system including the extended overlap LGBP for feature extraction and the FA variant for feature selection. Section IV presents evaluation of the proposed system in comparison with other search methods and related facial expression recognition research. Section V draws conclusions and identifies future directions.

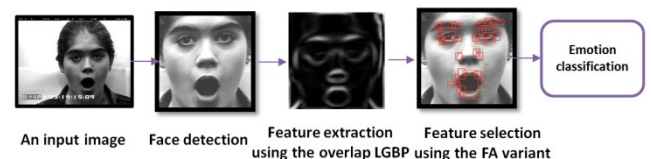


Fig. 1. System architecture of the proposed system

II. RELATED WORK

A. Feature Extraction

Feature extraction plays an important role in facial expression recognition applications. In general, the feature extraction algorithms can be categorised as geometric and appearance/texture models. Cootes et al. [9] proposed a well-known geometric feature extraction model, i.e. Active Shape Mode (ASM), to robustly locate and recognize the objects in the presence of noise, rotations and occlusions. Subsequently, Cootes et al. [10] extended ASM to Active Appearance Model (AAM), which extracts both geometric and appearance features from an input image. Their experimental results also indicated the effectiveness of the AAM, which outperformed ASM in various computer vision tasks. Also, Cristinacce and Cootes

[11] proposed a Constrained Local Model (CLM) to achieve efficient and robust real-time facial landmark detection. Although CLM does not deal with appearance feature extraction, it achieves impressive accuracy for face tracking and facial component detection for diverse real-life challenging situations.

Texture-based feature extraction algorithms are very popular for facial expression recognition applications. The widely used texture feature extraction algorithms include Local Binary Patterns (LBP) [12], Gabor filter [13], and Scale Invariant Feature Transform (SIFT) [14]. These models show very promising performances but most of them indicate high computational complexity. However, LBP proposed by Ojala et al. [12] is well-known for computational simplicity and robustness to illumination changes. LBP shows limitations in dealing with large-scale textural structures due to the fixed 3×3 pixel circular neighborhood [12]. To extract the large-scale textural structures, Ojala et al. 2002 [15] further extended the LBP model to support different sizes of circular neighborhood. Rotation invariant features were also introduced to deal with the rotation variation problems in their work. Zhang et al. [16] proposed the LGBP algorithm to retrieve more efficient discriminative features, which first applies Gabor filter to produce magnitude image and then applies the LBP operator to generate the texture description of the input image. The LGBP model proves to be more efficient in comparison with the LBP operator. However they both generate high dimensional features.

B. Feature Selection

Feature selection algorithms are usually used to reduce feature dimension by identifying the most discriminative features and removing the redundant ones. The commonly used feature dimension reduction algorithms include Principal Component Analysis (PCA) [17], Linear Discriminant Analysis (LDA) [18], and Independent Component Analysis (ICA). PCA is one of the most commonly used feature dimensionality reduction algorithms in face recognition, however experimental results indicated that it lacks discriminative power [18, 19]. The boosted-LBP algorithm was proposed by Shan et al. [18], which applied Ada-Boosting for feature selection by identifying optimal LBP generated sub-regions instead of selecting histograms. The boosted-LBP showed high recognition accuracy when evaluated with the CK+ database.

In recent years, evolutionary optimization algorithms have attracted significant attention and have been used extensively for feature optimization problems [19, 20, 21 and 22]. For instance, GA and PSO are the most commonly used evolutionary search methods for feature selection [19, 20]. As an example, Sun et al. [22] applied Genetic Algorithm (GA) to identify optimal features for 3D face recognition whereas Particle Swarm Optimization (PSO) has also been applied to identify discriminative motion-related bodily features for the regression of valence and arousal dimensions for bodily expression recognition [20]. Mistry et al. [23] have also applied micro-GA embedded with PSO for feature optimization for facial expression recognition.

III. THE PROPOSED FACIAL EXPRESSION RECOGNITION SYSTEM

We introduce the proposed facial expression recognition system in detail in this section, which consists of three key steps, i.e. a modified LGBP-based feature extraction, a modified FA-based feature optimization and emotion recognition. Each key step is introduced in detail in the following.

A. Feature Extraction Using Overlap LGBP

First of all, pre-processing is applied in order to reduce image noise. A histogram equalization method is first used to improve the contrast of an input image. A bilateral filter is then applied to reduce image noise while preserving the edges. We subsequently apply Viola and Jones's face detection algorithm [24] provided in the OpenCV package to detect the face region of the input image.

In order to deal with illumination changes and pose variations, we propose an extended overlap LGBP for facial feature extraction in this research. As discussed earlier, the original LGBP operator is the combination of LBP and Gabor filters. The 2D Gabor filter is first applied to an input image to decompose a face image and then the LBP operator is applied to generate texture description. Since the original LBP operator uses a circular neighborhood for texture description, it is likely to lose important information while transiting between sub-regions [18, 23]. Therefore, we propose a modified extended overlap LGBP for feature extraction. This proposed LGBP operator overlaps the last column of the first LBP sub-region with the first column of the neighboring LBP sub-region. In this way, this overlap process enables the retrieval of any missing information from the corners of the sub-regions to provide a more refined facial representation. The empirical results indicate that this proposed LGBP operator possesses more discriminative power in comparison to the original LGBP. The feature extraction process using the proposed overlap LGBP is illustrated in Figure 2.



Fig. 2. The proposed extended overlap LGBP

In this research, we have used a 3×3 pixel circular neighborhood for each sub-region. The face region image retrieved from face detection process has a size of 75×75 pixels. Therefore, we have obtained 676 sub-regions after applying the proposed extended overlap LGBP operator.

However, both the original and proposed LGBP operators have the disadvantage of high dimensionality, which makes them less efficient in real-time applications [19, 25]. In order to deal with this issue, in this research, we propose a FA variant to conduct feature dimensionality reduction and identify the

most discriminative feature subsets to inform subsequent facial expression recognition.

B. Feature Selection using the FA Variant

In this section, we introduce the proposed FA variant in detail. FA is originally proposed by Yang [26], which is a swarm intelligence algorithm inspired by natural behaviours of fireflies. In FA, a firefly with less brightness is attracted towards other fireflies with more brightness. The algorithm employs the attraction and attractiveness behaviours of the fireflies to explore the search space and identify optimal solutions. FA has efficient local and global search strategies where brightness is linked with objective functions. The following three basic idealised rules are followed by artificial fireflies to conduct search in the search space, i.e. (1) All fireflies are unisex and are attracted towards brighter ones regardless of their sex. (2) The attractiveness of firefly is directly proportional to the brightness of the firefly. Therefore, the less bright firefly will be attracted towards a brighter one. (3) The fitness function defines the brightness (light intensity) of a firefly.

Moreover, in FA, the attractiveness of the firefly varies with the distance between two fireflies, i.e. attractiveness is decreased along with the increase of the distance. However, the brightest firefly moves randomly in the search space since there is no any other more attractive firefly with stronger light intensity.

FA formulates two important aspects, i.e. the variation of light intensity and attractiveness. The light intensity $I(r)$ decreases as the distance increases from its source while the media also absorbs light. The light intensity varies with the distance which is defined by the following equation.

$$I(r) = I_0 \exp(-\gamma r^2) \quad (1)$$

where r denotes the distance between two fireflies. I_0 is initial light intensity at $r = 0$, and γ denotes the light absorption coefficient constant.

Let us consider the attractiveness $\beta(r)$ of a firefly is proportional to a firefly's brightness. It also varies with the distance r_{ij} between two fireflies i and j . The attractiveness function of a firefly in FA is denoted as follows.

$$\beta(r) = \beta_0 \exp(-\gamma r^2) \quad (2)$$

where r denotes the distance between two fireflies while β_0 is initial attractiveness at $r = 0$, and γ denotes the light absorption coefficient.

The following equation is also used to calculate the distance between fireflies i and j at x_i and x_j .

$$r_{ij} = \|x_i - x_j\| = \sqrt{\sum_{k=1}^d (x_{i,k} - x_{j,k})^2} \quad (3)$$

where d denotes the dimension index of the given problem.

The position updating of a less bright firefly towards a brighter one is shown in Equation (4).

$$x_i = x_i + \beta_0 e^{-\gamma r_{ij}^2} (x_i - x_j) + \alpha [\text{rand} - \frac{1}{2}] \quad (4)$$

where the second term indicates the effects of the attractiveness of fireflies while the third term α is a randomization parameter and rand is random number generator distributed in the range of $[0, 1]$ [26].

The FA mechanism is very efficient in local search but shows limitations for global exploration [26]. Therefore, three mutation operators, i.e. Gaussian, Cauchy and Levy distributions [27], are applied in this research to balance the local and global exploration of the conventional FA to mitigate premature convergence. Specifically, the mutation techniques are applied to the global best solution identified by the FA to enable long jumps to avoid local optimum traps. The three mutation operators, i.e. Gaussian, Cauchy and Levy distributions, are employed since they are nominated well-accepted evolutionary mutation operators. First of all, the global best solution identified by the FA is further enhanced by the Gaussian mutation, which is defined in the following equation.

$$P_{g1,s}(d) = P_{g,s}(d) + (X_d^{\max} - X_d^{\min}) \times \phi(o, h) \quad (5)$$

where $\phi(o, h)$ indicates the Gaussian distribution and o represents the mean or expectation of the distribution with h as the standard deviation. X_d^{\max} and X_d^{\min} indicate the upper and lower bounds of decision vectors in the d^{th} dimension respectively with $d = 1, 2, \dots, n$. The newly generated global best solution $P_{g1,s}$ using the Gaussian distribution is used to replace the previous global best $P_{g,s}$ if $P_{g1,s}$ has better fitness than $P_{g,s}$.

Afterwards, the Cauchy mutation operator is applied to further improve the global best, which is defined in Equation (6).

$$P_{g2,s}(d) = P_{g,s}(d) + (X_d^{\max} - X_d^{\min}) \times \psi(g, s) \quad (6)$$

where $\psi(g, s)$ indicates the Cauchy mutation and g and s respectively represent the location parameter (indicating the location of the peak) and the scale parameter of this distribution. The scale parameter, s , also decreases linearly during the search. The newly generated global best solution $P_{g2,s}$ is used to replace the previous global best $P_{g,s}$ if $P_{g2,s}$ has better fitness than $P_{g,s}$.

Finally, we employ Levy distribution to further increase the exploration and exploitation capabilities of the global best. This process is defined in Equations (7) and (8).

$$P_{g3,s}(d) = P_{g,s}(d) + (X_d^{\max} - X_d^{\min}) \times L(\mu, k, \eta) \quad (7)$$

$$L(\mu, k, \eta) = e^{-\mu|k|^\eta} \quad (8)$$

where μ is the scale factor ranging from -1 to 1 and η is the levy's index with the value ranging from 0 to 2. The newly

generated global best solution $P_{g_{3,s}}$ is used to replace the previous global best $P_{g,s}$ if $P_{g_{3,s}}$ has better fitness than $P_{g,s}$.

The fitness function defined to evaluate each firefly consists of two main criteria, i.e. the number of selected features and classification accuracy. The classification accuracy in fitness function is the accuracy for each individual emotion class rather than a combined overall accuracy for all emotions.

$$F(x) = w_a * acc_x + w_f * (number_feature_x)^{-1} \quad (9)$$

where w_a and w_f are two predefined constant weights for classification accuracy and the number of selected features, respectively, with $w_a = 1 - w_f$. In this research, w_a is set to 0.9 and w_f is set to 0.1 since we consider classification accuracy is more important than the number of selected features. This FA variant is applied to each expression category to identify its discriminative features to inform emotion recognition.

C. Emotion Classification

In this work, diverse classifiers have been employed to detect the seven emotions (i.e. anger, happiness, sadness, surprise, disgust, fear, and neutral). NN, multi-class SVM, and the SVM-based and NN-based ensembles with SVM and NN as base classifiers respectively are employed to conduct emotion classification. The feature subsets retrieved by the FA-based feature selection are used as the inputs to the classifiers. The input layer for NN is set to 50-65 nodes, where each node indicates an optimized feature recommended by the modified FA algorithm. The NN classifier has one hidden layer and one output layer with seven nodes representing each emotion respectively. The optimal setting of NN is identified using a trial-and-error method. Moreover, the grid-search method is also employed to obtain the optimal parameter settings for the SVM classifier in order to achieve optimal performance.

Two ensemble classifiers, i.e. the NN-based and SVM-based ensembles, are also employed to improve classification accuracy. The optimal settings obtained for each single model NN and SVM classifier mentioned above are also applied to the setting of each base classifier within each ensemble. Both ensembles employ three base classifiers and use a weighted majority voting combination method to produce final classification.

IV. EVALUATION

The proposed system has been evaluated using within and cross database evaluation with images extracted from CK+, MMI and JAFFE databases. First of all, we compare the proposed extended overlap LGBP operator with the original LGBP. Then the proposed FA-based feature selection algorithm is evaluated against the classical search algorithms such as FA, GA and PSO. Single and ensemble classifiers, such as NN, SVM, and NN-based and SVM-based ensembles, are used for the classification of seven emotions. For all the experiments, a set of 250 images from CK+ is employed for training while a set of 175 images from each of three databases (i.e. CK+, JAFFE and MMI) is used for testing.

First, the optimal parameter settings of the proposed FA and other search methods are identified. The parameter settings for the proposed FA are as follows: population size = 30, initial attractiveness = 1.0, randomisation parameter = 0.2, absorption coefficient = 1.0 and maximum iterations = 500. The original FA has also adopted the above configuration. The following settings are applied to PSO: maximum velocity = 0.6, inertia weights = 0.78, population size = 30, acceleration constants $c_1=c_2=1.2$ and maximum iterations = 500. The GA employs the settings of crossover probability = 0.6, mutation probability = 0.05, and maximum generations = 500.

Since the proposed FA variant, FA, PSO and GA are evolutionary algorithms, due to their randomization characteristics, we perform 30 benchmark runs for each of these algorithms in order to conduct a fair comparison. The average accuracy of the 30 runs of each model is employed for performance comparison. The first experiment conducted uses 250 and 175 images from CK+ for training and testing, respectively. Table I shows the results of the proposed system with the first three rows indicating the top three trials out of 30 runs, where the proposed LGBP and FA variant are combined with diverse classifiers. In order to conduct the comparison between the proposed overlap and the original LGBP, the bottom two rows in Table I present the results obtained using all the raw features extracted by the proposed extended overlap LGBP and original LGBP without any feature selection process.

TABLE I. EXAMPLE CLASSIFICATION RESULTS FOR THE PROPOSED SYSTEM AND THOSE OBTAINED USING ALL THE RAW FEATURES EXTRACTED BY THE PROPOSED AND THE ORIGINAL LGBP

	Number of selected features	NN (%)	SVM (%)	NN-based Ensemble (%)	SVM-based Ensemble (%)
1	60	90.00	90.50	96.11	97.22
2	65	91.15	92.00	97.32	98.50
3	70	89.90	89.43	93.45	93.40
4	676 (the proposed LGBP)	70.33	71.21	78.70	79.32
5	625 (the original LGBP)	63.20	64.83	70.00	71.20

TABLE II. THE AVERAGE CLASSIFICATION ACCURACY OVER 30 RUNS FOR ALL THE FEATURE SELECTION ALGORITHMS FOR WITHIN DATABASE EVALUATION (CK+)

	Number of selected features	NN % (30 Runs)	SVM % (30 Runs)	NN-based Ensemble % (30 Runs)	SVM-based Ensemble % (30 Runs)
GA	100-200	74.60	76.90	78.88	80.00
PSO	110-200	76.33	78.70	81.33	82.50
FA	60-90	78.45	79.00	84.67	85.40
Prop. FA	50-65	94.55	94.90	97.33	98.45

For each algorithm, the results obtained by the SVM-based ensemble classifier outperform those of all the other classifiers. The highest accuracy achieved by the proposed system with the overlap LGBP and the FA variant is 98.5% with 65 features selected when combined with the SVM-based ensemble. The system with FA-based feature selection outperforms those without any feature selection process shown in the last two

rows in Table I, significantly. Also the results gained using the proposed overlap LGBP operator show significant improvement over those obtained using the original LGBP, which proves the efficiency of the proposed LGBP-based feature extraction algorithm.

Moreover, we have also compared the performances of the FA variant with those of FA, GA and PSO in Table II, where all the algorithms are trained and tested with 250 and 175 images from CK+. A set of 30 runs is performed for each optimization algorithm. For all the experiments, all search algorithms in combination with the SVM-based ensemble achieve the highest accuracy. As shown in Table II, the FA variant outperforms FA, GA and PSO. Also the FA variant is able to reduce the number of features to the range of 50-65, which is significantly lower than the number of features retrieved by conventional FA (60-90), PSO (110-200) and GA (100-200).

In order to further prove the efficiency of the proposed system, we have conducted cross-database evaluation with 250 images from CK+ for training and a set of 175 images from JAFFE and MMI for testing, respectively. A set of 30 runs is also conducted for each optimization algorithm. Table III shows the results for cross-database evaluation, where 175 images from JAFFE database are used for testing.

TABLE III. THE AVERAGE CLASSIFICATION ACCURACY OVER 30 RUNS FOR ALL THE OPTIMIZATION ALGORITHMS FOR CROSS-DATABASE EVALUATION (CK+ FOR TRAINING AND JAFFE FOR TESTING)

	Number of selected features	NN % (30 Runs)	SVM % (30 Runs)	NN-based Ensemble % (30 Runs)	SVM-based Ensemble % (30 Runs)
GA	100-200	72.21	73.65	77.00	78.30
PSO	110-200	73.13	75.90	78.72	79.89
FA	60-90	76.35	78.80	82.43	82.95
Prop. FA	50-65	79.45	80.00	85.40	87.75

TABLE IV. THE AVERAGE CLASSIFICATION ACCURACY OVER 30 RUNS FOR ALL THE SELECTED FEATURE SELECTION ALGORITHMS FOR CROSS-DATABASE EVALUATION (CK+ FOR TRAINING AND MMI FOR TESTING)

	Number of features	NN % (30 Runs)	SVM % (30 Runs)	NN-based Ensemble % (30 Runs)	SVM-based Ensemble % (30 Runs)
GA	100-200	69.60	69.79	75.34	77.51
PSO	110-200	71.03	73.11	77.51	78.06
FA	60-90	74.09	76.23	81.86	84.24
Prop. FA	50-65	79.00	82.10	87.33	88.00

As shown in Table III for cross-database evaluation, the FA variant outperforms the other three search algorithms when integrated with each of the classifiers. Also, when SVM-based ensemble is used, the FA variant achieves the highest accuracy rate of 87.75% and outperforms FA, GA and PSO by 4.8%, 9.45% and 7.86%, respectively.

Moreover, another cross-database evaluation is also conducted with 250 images from CK+ for training and 175 images from MMI for testing. A set of 30 runs is also conducted for each optimization method. This experiment is

proven to be comparatively more challenging than other experiments. Evaluation results are illustrated in Table IV, which show more performance differentiation among the selected optimization algorithms.

As illustrated in Tables IV, trained with CK+ and tested upon MMI, the FA variant achieves the highest accuracy of 88% and outperforms conventional FA, GA and PSO by 3.76%, 10.49% and 9.94%, when SVM-based ensemble is applied.

TABLE V. COMPARISON BETWEEN THE PROPOSED FA AND NON-EVOLUTIONARY FEATURE SELECTION METHODS

	Average number of selected features	CK+ SVM-based Ensemble %	JAFFE SVM-based Ensemble%	MMI SVM-based Ensemble %
PCA	250	81.50	77.00	76.95
ICA	250	81.00	77.45	77.20
LDA	200	82.75	78.80	77.55
Prop. FA	58	98.45	87.75	88.00

Besides empirical comparison against evolutionary feature optimisation algorithms, we further compare the proposed FA with non-evolutionary feature selection methods such as PCA, ICA, and LDA. In each experiment, we apply the proposed LBP operator to extract the initial features and then apply above-mentioned non-evolutionary algorithms for feature selection. Table V shows the detailed comparison between the proposed FA and non-evolutionary feature selection methods. As illustrated in Table V, the proposed FA outperforms PCA, ICA, and LDA by a significant margin when tested on all three datasets.

Moreover, to show the clear overview of the achieved performances over inter-dataset and cross-dataset evaluation, we have calculated the variance for each of the experiment. Table VI shows the variance for all the methods when evaluated using the above three datasets and the SVM-based ensemble classifier.

TABLE VI. PERFORMANCE VARIANCE FOR EACH METHOD

	CK+ SVM-based Ensemble (30 Runs)	JAFFE SVM-based Ensemble (30 Runs)	MMI SVM-based Ensemble (30 Runs)
GA	11.53	21.64	12.24
PSO	21.66	12.60	9.62
FA	21.86	5.64	9.99
Prop. FA	3.49	4.26	5.09

TABLE VII. COMPARISON WITH RELATED RESEARCH FOR CK+

Methods	Methodology	Classes	Evaluation Strategy	Recognition Rate (%)
Shan et al. [18]	Boosted LBP+SVM	7	10-fold	91.40
Zhong et al. [28]	CSPL	6	10-fold	89.89
This research	Overlap LBP +FA variant+ Ensemble (SVM)	7	Average of 30 runs with 46.6% for testing	98.45
This research	Overlap LBP +FA variant+ Ensemble (SVM)	7	10-fold	98.15

The proposed system is also compared with other existing state-of-the-art facial expression recognition research. Tables VII & VIII show the detailed comparison between the proposed system and other related research for CK+ and MMI databases, respectively.

TABLE VIII. COMPARISON WITH RELATED RESEARCH FOR MMI

Methods	Methodology	Classes	Evaluation Strategy	Recognition Rate (%)
Elaiwat et al. [29]	Spatio-temporal RBM based model	6	10-fold	81.63 (trained with CK+)
Zhong et al. [28]	CSPL	6	10-fold	73.53 (trained with CK+)
This research	Overlap LBP +FA variant+ Ensemble (SVM)	7	Average of 30 runs	88.00 (trained with CK+)
This research	Overlap LBP +FA variant+ Ensemble (SVM)	7	10-fold	79.85 (trained with CK+)

As indicated in Table VII, the proposed system outperforms all other related research when using CK+ for training and testing. Also, as shown in Table VIII, when using CK+ for training and MMI for testing, the proposed system outperforms the work of Elaiwat et al. [29] and Zhong et al. [28] significantly. The proposed LGBP-based feature extraction and the FA variant based feature optimization account for the great efficiency and robustness of the proposed system.

Furthermore, we compare the computational efficiency of our algorithm with all other evolutionary and non-evolutionary feature selection algorithms in Table IX. The computational cost shown in Table IX includes the execution of the proposed LBP for feature extraction, the corresponding method for feature selection and the SVM-based ensemble for classification.

TABLE IX. COMPUTATIONAL COST OF THE PROPOSED FA VARIANT AND OTHER BASELINE METHODS

	Average number of selected features	Computational Cost (milliseconds)
GA	150	381
PSO	155	365
FA	75	270
PCA	250	410
ICA	250	400
LDA	200	395
Prop. FA	58	255

As illustrated in Table IX, the computational cost of each method depends on the number of selected features for classification. A fewer number of selected features will result in comparatively lower computational cost and vice versa. Since the propose FA has the smallest number of selected features, it has more optimal computational cost.

V. CONCLUSION

In this research, we have proposed a facial expression recognition system with the proposed overlap LGBP operator for feature extraction and a FA variant for feature optimization. Diverse classifiers are used to conduct the recognition of the

seven facial expressions. The proposed FA variant identifies the least number of features and outperforms other conventional search methods such as FA, GA and PSO, significantly. The proposed system achieves an average accuracy of 98.45% over 30 runs when evaluated with CK+ database images. The system also shows promising performance for cross-database evaluation and achieves an average accuracy of 87.75% for JAFFE and 88% for MMI over 30 runs respectively. It also outperforms state-of-the-art related facial expression recognition research significantly. In future work, other hybrid or multi-objective FAs [30] will also be explored to solve optimization problems with multiple criteria.

REFERENCES

- [1] L. Zhang, B. Fielding, P. Kinghorn, K. Mistry. "A Vision Enriched Intelligent Agent with Image Description Generation", In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, Singapore. 2016, pp. 1488–1490.
- [2] A.S. G'Mussel and J. Hewig. 2013. "The value of a smile: Facial expression affects ultimatum-game responses". *Judgment and Decision Making*, 8 (3), 381-385.
- [3] L. Zhang, M. Jiang, D. Farid, A. Hossain. 2013. "Intelligent Facial Emotion Recognition and Semantic-based Topic Detection for a Humanoid Robot". *Expert Systems with Applications*, 40 (2013), pp. 5160-5168.
- [4] L. Zhang, K. Mistry, M. Jiang, S.C. Neoh, and M.A. Hossain. "Adaptive facial point detection and emotion recognition for a humanoid robot". *Computer Vision and Image Understanding*, 140(2015): 93–114.
- [5] E. Vural, M. Cetin, A. Ercil, G. Littlewort, M. Bartlett and J. Movellan, 2008. "Automated Drowsiness Detection for Improved Driver Safety". In *Proceedings of the International Conference on Automotive Technologies*, 2008.
- [6] P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. "The extended Cohn-Kanade dataset (CK+): A complete expression dataset for action unit and emotion-specified expression", in *Proceedings of the 3rd International Workshop on CVPR for Human Communicative Behavior Analysis*, USA, 2010, pp. 94-101.
- [7] M.J. Lyons, S. Akemastu, M. Kamachi and J. Gyoba. "Coding Facial Expressions with Gabor Wavelets". In *Proceedings of 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 200-205 (1998).
- [8] M. Pantic, M.F. Valstar, R. Rademaker, and L. Maat. (2005). "Web-based database for facial expression analysis", in *Proceedings of IEEE Int'l Conf. Multimedia and Expo*, The Netherlands, 317 – 321.
- [9] T.F. Cootes, C.J. Taylor, D. J. Cooper and J. Graham, "Active Shape Models-Their training and Applications", in *Computer Vision and Image Understanding*, Vol. 61, pp. 38-59, 1995.
- [10] T.F. Cootes, G.J. Edwards and C.J. Taylor, "Active Appearance Models", In *Proc. European Conference on Computer Vision*, (H.Burkhardt & B. Neumann Eds). Vol. 2, pp. 484-498, Springer. 1998.
- [11] D. Cristinacce and T. Cootes, "Feature detection and tracking with constrained local models". In *Proceedings of British Machine Vision Conference*, Vol. 3, BMVA Press, 2006, pp. 929–938.
- [12] T. Ojala, M. Pietikäinen and D. Harwood, "A comparative study of texture measures with classification based on featured distribution", *Pattern Recognition* 29 (1) (1996) 51–59.
- [13] M. Valstar and M. Pantic. "Fully automatic facial action unit detection and temporal analysis", in *Proceedings of IEEE Conference on computer vision and pattern recognition (CVPR)*, 2012, pp. 149-156.
- [14] L.G. David. "Object recognition from local scale invariant features", *the international conference on computer vision*, 1999, pp. 1150-1157.
- [15] T. Ojala, M. Pietikäinen and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (7) (2002) 971–987.

- [16] W. Zhang, S. Shan, X. Chen and W. Gao, "Local Gabor binary patterns based on mutual information for face recognition", *Int. J. Image Graph.* 7 (3) (2007) 1-17
- [17] K. Pearson. "On lines and planes of closest fit to system of points in space", *Philosophical magazine* 2 (11) (1901), 559-572.
- [18] C. Shan, S. Gong, P.W. McOwan. "Facial expression recognition based on local binary patterns: a comprehensive study", *Image Vis. Comput.* 27 (2009) 803-816.
- [19] S.C. Neoh, L. Zhang, M. Mistry, M.A. Hossain, C.P. Lim, N. Aslam, and P. Kinghorn. "Intelligent Facial Emotion Recognition Using a Layered Encoding Cascade Optimization Model". *Appl Soft Comput.* Volume 34, 2015, 72-93.
- [20] Y. Zhang, L. Zhang, S.C. Neoh, K. Mistry, and M.A. Hossain. "Intelligent affect regression for bodily expressions using hybrid particle swarm optimization and adaptive ensembles". *Expert Syst. Appl.* 42 (2015): 8678-8697.
- [21] Y. Zhang, L. Zhang, and M.A. Hossain. (2015). "Adaptive 3D facial action intensity estimation and emotion recognition", *Expert Syst. Appl.* 42 (2015): 1446-1464.
- [22] Sun, Y. and Yin, L. "A Genetic Algorithm Based Approach for 3D Face Recognition". In *3D Imaging for Safety and Security*, Koschan, Andreas, Pollefeys, Marc, Abidi, Mongi (Eds.). 2007. pp. 95-118.
- [23] K. Mistry, L. Zhang, S.C. Neoh, C.P. Lim. and B. Fielding. 2016. "A micro-GA Embedded PSO Feature Selection Approach to Intelligent Facial Emotion Recognition". *IEEE Transactions on Cybernetics*. doi: 10.1109/TCYB.2016.2549639
- [24] P.A. Viola and M.J. Jones. "Rapid Object Detection using a Boosted Cascade of Simple Features". *CVPR* . 511-518. 2001.
- [25] K. Mistry, L. Zhang, and J. Barnden. "Intelligent facial expression recognition with adaptive feature extraction for humanoid robot", In *Proceedings of International Joint Conference on Neural Networks (IJCNN)*. Ireland. 2015.
- [26] X.S. Yang. 2010. "Firefly algorithm, Levy's flight and global optimization", *Research and Development in intelligent systems* 26 (2010) 209-218.
- [27] X.S. Yang. "Nature Inspired Metaheuristic Algorithms". Luniver press, 2008.
- [28] L. Zhong, Q. Liu, P. Yang, B. Liu, J. Huang and D. Metaxas. "Learning active facial patches for expression analysis", In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2562-2569. 2012.
- [29] S. Elaiwat, M. Bennamoun and F. Boussaid. "A spatio-temporal RBM-based model for facial expression recognition", *Pattern Recognition*, 49 (2016), pp. 152-161.
- [30] L. Zhang, K. Mistry, S.C. Neoh, and C.P. Lim. (2016). "Intelligent facial emotion recognition using moth-firefly optimization". *Knowledge-Based Systems*. Volume 111, 2016, pp. 248-267.